

英語学習者の文発声における韻律自動評定

加藤圭介† 野沢和典‡ 山下洋一†

†立命館大学理工学部情報学科

‡立命館大学経済学部

〒525-8557 滋賀県草津市野路東 1-1-1

E-mail: †(kato,yama)@slp.cs.ritsumeikan.ac.jp, ‡nozawa@ec.ritsumeikan.ac.jp

概要 本報告では、日本人英語学習者の英文発話における韻律を自動評定する手法について述べる。学習者と英語母語話者の発話を比較し、基本周波数、パワー、発話長の類似度を算出して韻律パラメータとする。2つの発話を比較する際には単語や単語境界部など様々な比較単位ごとに比較し、比較単位による結果の違いを考察した。また、基本周波数とパワーに関してはパターン距離などの従来手法に加え、回帰曲線近似誤差を用いた評定手法を提案し、評定結果の妥当性を検証した。さらに、複数の韻律パラメータを組み合わせ、学習者発話の韻律を評定するモデルを作成した。

キーワード : CALL、韻律、発話比較、重回帰モデル

Prosodic Scoring for Sentence Speech Uttered by English Learners

Keisuke Kato†, Kazunori Nozawa‡ and Yoichi Yamashita†

†Department of Computer Science, Ritsumeikan University

‡Faculty of Economics, Ritsumeikan University

1-1-1 Noji-Higashi, Kusatsu-shi, Shiga, 525-8557 Japan

E-mail: †(kato,yama)@slp.cs.ritsumeikan.ac.jp, ‡nozawa@ec.ritsumeikan.ac.jp

Abstract In this paper, we describe techniques to score prosody of sentence speech uttered by English learners. Based on the comparison between learners' speech and native speaker's speech, we make prosodic parameters by calculating the similarity of learners' speech and native speaker's speech about F0, power and duration. The comparison is carried out every comparison unit, such as a word, a word boundary, and so on. We try a new scoring measure in terms of the approximate error of regression fitting, as well as pattern distance, for F0 and power. Moreover, we make a multiple regression model for scoring prosody of English learners' speech by combining two or more prosodic parameters.

key words : CALL, prosody, speech comparison, multiple regression model

1. はじめに

日本人が英語を上手く話せない理由として、学校における英語教育の問題や日本語と英語の言語的差異が挙げられている[1]。特に韻律に関しては、発音と比べて教育を受ける機会が少なく、また、日本語と英語のアクセント生成法の違いのため、日本人にとって習得は容易ではない。そのような日本人の英語発声練習を支援する方法として、計算機を利用した語学教育(CALL: Computer Assisted Language Learning)が注目されており、音声情報処理技術を用いた CALL システム開発に関する研究が盛んに行われている。

本研究では、日本人英語学習者(以下、学習者)の英文発話における韻律を評定する手法を検討している。英語母語話者(以下、ネイティブ)発話と学習者発話の基本周波数、パワー、発話長を様々な手法で比較し、計算機による評定値(以下、自動評定値)を算出する。英語教師による評定値(以下、教師評定値)と強い相関を持つような自動評定値を算出することを目的とする。また、複数の韻律パラメータを組み合わせることで学習者発話の韻律を評定するモデルを作成する。

2. 韻律自動評定手法

本研究では、学習者が発声する英文は既知であり、同じ英文をネイティブ話者が発声したデータを利用可能である状況を想定している。そこで、ネイティブと学習者との発話比較による評定手法を検討する。

2つの発話を比較して評定する際には、線形伸縮やDP マッチングによってネイティブ発話と学習者発話の対応を取り、発話全体における基本周波数などの誤差を算出し自動評定値とする手法が提案されている[2]。しかし、評定対象を日本人とする場合、発話中に現れる日本人の英文発声特徴を考慮することで自動評定値と教師評定値との相関を強めることができると考えられる。そこで、本研究では日本人の英文発話の良し悪しが顕著に表れる部分について検討し、それらを比較単位として、評定を行う[3]。

両者の発話について、パラメータ(基本周波数、パワー、発話長)を比較単位ごとに比較し、その平均誤差を評定値とする。比較単位数を I 、ネイティブと学習者の $i(i=1 \dots I)$ 番目の比較単位におけるパラメータをそれぞれ $P_N(i)$ 、 $P_L(i)$ としたとき自動評定値 S を次式のように定義する。

$$S = \frac{1}{I} \sum_{i=1}^I |P_N(i) - P_L(i)| \quad (1)$$

なお、比較単位とは単語、単語境界部(後述)、発話全体など発話の比較に用いる基本単位を指し、英語・日本語音素混合モデルを用いた自動ラベリングの結果を基にその区間を決定する[4]。また、基本周波数とパワーはフレーム周期 5ms、フレーム長 20ms で特徴抽出し、対数化した後、平均値を 100 に揃える。基本周波数については無声音区間の線形補間とスムージング処理を行う。次章で、基本周波数、パワー、発話長に基づく韻律パラメータと比較単位について述べる。

3. 韻律パラメータ

3.1. 基本周波数

3.1.1. 比較単位

基本周波数に関する韻律パラメータの比較単位について述べる。杉藤は語群の発話について、図1に模式的に示すように、ネイティブ発話の基本周波数パターンには単語と単語の境界(以下、単語境界部)において平板な形が観測されるが、日本人には観測されないと指摘している[5]。そこで、発話の上手さは単語境界部に顕著に現れると考え、比較単位として単語境界部を用いる。単語境界部については、以下の2つの区間を検討する。

- 単語と単語の境界における前後の英語音節を合わせた区間(図2)(以下、単語境界シラブル部)
- 単語と単語の境界における前後の日本語音節(単語を日本語読みした時の音節)を合わせた区間(図3)(以下、単語境界モーラ部)

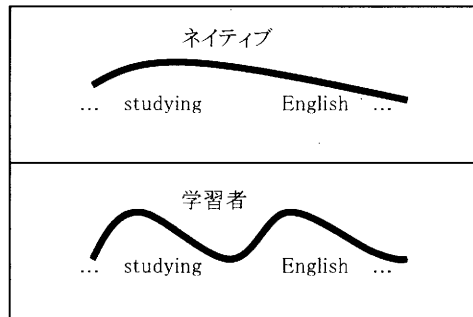


図1 ネイティブ・学習者発話の単語境界部における基本周波数パターンの例

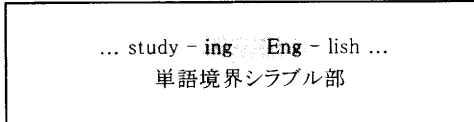


図2 単語境界シラブル部の例

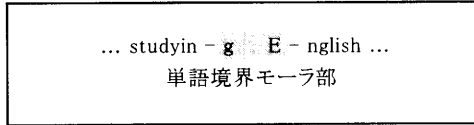


図3 単語境界モーラ部の例

3.1.2. 基本周波数パターン距離

「基本周波数パターン距離」とは、比較単位内で基本周波数の平均を揃えたネイティブ発話と学習者発話のフレーム間距離のことである。なお、ネイティブ発話と学習者発話の時間長が異なる場合には、音節内での線形伸縮によってフレーム間の対応付けを行う。 $i(i=1...I)$ 番目の比較単位において、ネイティブ発話と学習者発話の基本周波数の平均値をそれぞれ $\overline{f_N(i)}$ 、 $\overline{f_L(i)}$ とする。また、ネイティブ発話の $k(k=1...K)$ 番目のフレームにおける基本周波数を $f_N(k)$ とし、これに対応する学習者発話の基本周波数を $f_L(k)$ とするとき、 i 番目の比較単位における基本周波数パターン距離 $f_pd(i)$ を次のように定義する。

$$f_pd(i) = \text{avg} \left(\left| f_N(k) - f_L(k) - \left(\overline{f_N(i)} - \overline{f_L(i)} \right) \right| \right) \quad (2)$$

このとき、基本周波数パターン距離 F_PD を次のように定義する。

$$F_PD = \text{avg} (f_pd(i)) \quad (3)$$

3.1.1 節で述べた「単語境界シラブル部」、「単語境界モーラ部」の他に「発話全体」と「単語」も比較単位として用い、結果の違いを考察する。

3.1.3. 基本周波数パターン回帰直線の傾き差

「基本周波数パターン回帰直線の傾き差」とは、比較単位におけるネイティブ発話と学習者発話の回帰直線の傾きを比較したものである。 $i(i=1...I)$ 番目の比較単位におけるネイティブ発話と学習者発話の基本周波数パターン回帰直線の傾きをそれぞれ $\Delta f_N(i)$ 、

$\Delta f_L(i)$ とする。このとき基本周波数パターン回帰直線の傾き差 F_SD を次のように定義する。

$$F_SD = \text{avg} \left(\left| \Delta f_N(i) - \Delta f_L(i) \right| \right) \quad (4)$$

図4に単語を比較単位とした基本周波数パターン回帰直線の傾きの例を示す。

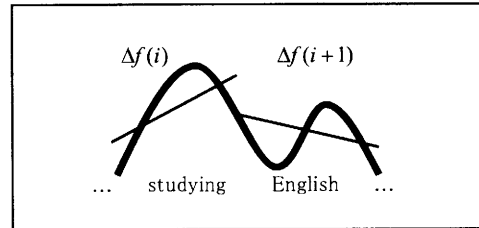


図4 基本周波数パターン回帰直線の傾き (比較単位は単語)

3.1.1 節で述べた「単語境界シラブル部」、「単語境界モーラ部」の他に「単語」も比較単位として用い、結果の違いを考察する。

3.1.4. 基本周波数パターン回帰曲線近似誤差の差

学習者が英文を発声する際には、各単語の発音を強く意識しすぎると英語教師が指摘している。単語の発音を意識しすぎると、各単語に固有の韻律の特徴が英文発話中にそのまま現れてしまい、単語単位で見ると基本周波数の上下変化が大きくなっていると考えられる。また逆に、学習者発話はイントネーションの変化幅が小さく抑揚が少ないとも言われ[6]、これは実験によっても示されている[7]。このことから、学習者発話の基本周波数パターンを回帰曲線で近似した場合の近似誤差は、前者の場合にはネイティブ発話よりも大きく、後者の場合には小さくなると予想できる。そこで、「基本周波数パターン回帰曲線近似誤差の差」を用いた評定手法を検討する。

「基本周波数パターン回帰曲線近似誤差の差」とは、ネイティブ発話と学習者発話について、比較単位における基本周波数パターンとその回帰曲線との2乗誤差(比較単位内のフレーム数で正規化)を比較したものである。 $i(i=1...I)$ 番目の比較単位におけるネイティブ発話と学習者発話の基本周波数パターン回帰曲線近似誤差をそれぞれ $fd_N(i)$ 、 $fd_L(i)$ とする。このとき基本周波数パターン回帰曲線近似誤差の差 F_ND を次のように定義する。

$$F_ND = \text{avg} \left(\left| fd_N(i) - fd_L(i) \right| \right) \quad (5)$$

なお、回帰曲線として回帰直線と回帰2次曲線を用いる。また、比較単位は「単語」、「単語境界シラブル部」、「発話全体」とする。図5に単語を比較単位とした場合の基本周波数パターン回帰曲線近似誤差の例を示す。

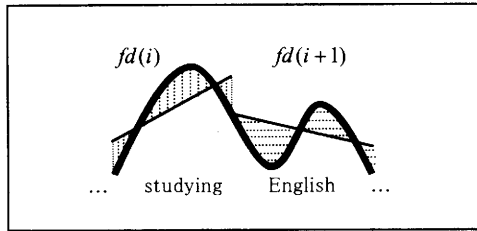


図5 基本周波数パターン回帰曲線近似誤差の例(比較単位は単語)

3.2. パワー

3.2.1. パワーパターン回帰曲線近似誤差の差

3.1.4 節で述べた理論は基本周波数だけでなくパワーに関しても当てはまると考えられる。そこで、基本周波数と同様に「パワーパターン回帰曲線近似誤差の差」を用いた評定手法を検討する。 $i(i=1\dots J)$ 番目の比較単位におけるネイティブ発話と学習者発話のパワーパターン回帰曲線近似誤差をそれぞれ $pd_N(i)$ 、 $pd_L(i)$ とする。このときパワーパターン回帰曲線近似誤差の差 P_ND を次のように定義する。

$$P_ND = \text{avg}(|pd_N(i) - pd_L(i)|) \quad (6)$$

なお、回帰曲線として回帰直線と回帰2次曲線を用い、比較単位は「単語」と「発話全体」とする。

3.3. 発話長

3.3.1. 総発話長の差

一般に、学習者の英文発話時間は不要な母音挿入や無音挿入のため、ネイティブよりも長くなる傾向がある。読み上げ音声の場合にはその傾向はより強くなると考えられる。そこで、ネイティブと学習者の総発話長の差を用いて評定を行う。ネイティブと学習者の総発話長をそれぞれ T_N 、 T_L とするとき、総発話長の差 T_D_{ALL} を次のように定義する。

$$T_D_{ALL} = |T_N - T_L| \quad (7)$$

3.3.2. 正規化単語長の差、正規化無音長の差

総発話長がネイティブと同程度であっても、発話リズムが悪ければ教師評定値は低くなると考えられる。そこで、正規化単語長の差、正規化無音長の差を用いて評定を行う。 $j(j=1\dots J)$ 番目の単語の発話長をそれぞれ $t_N(j)$ 、 $t_L(j)$ とするとき正規化単語長の差 T_RD_{WORD} を次のように定義する。

$$T_RD_{WORD} = \text{avg} \left(\left| \frac{t_N(j)}{T_N} - \frac{t_L(j)}{T_L} \right| \right) \quad (8)$$

また、 $j(j=1\dots J-1)$ 番目の単語に後続する無音の長さをそれぞれ $pau_N(j)$ 、 $pau_L(j)$ とするとき正規化無音長の差 T_RD_{PAUSE} を次のように定義する。

$$T_RD_{PAUSE} = \text{avg} \left(\left| \frac{pau_N(j)}{T_N} - \frac{pau_L(j)}{T_L} \right| \right) \quad (9)$$

3.3.3. 単語長の差、無音長の差

英語教師が学習者発話を評定する際には、発話時間とリズムの両方を合わせて評定すると考えられる。そこで、単語長の差、無音長の差を用いて評定を行う。ネイティブ発話と学習者発話における $j(j=1\dots J)$ 番目の単語の発話長をそれぞれ $t_N(j)$ 、 $t_L(j)$ とするとき単語長の差 T_D_{WORD} を次のように定義する。

$$T_D_{WORD} = \text{avg}(|t_N(j) - t_L(j)|) \quad (10)$$

また、 $j(j=1\dots J-1)$ 番目の単語に後続する無音の長さをそれぞれ $pau_N(j)$ 、 $pau_L(j)$ とするとき無音長の差 T_D_{PAUSE} を次のように定義する。

$$T_D_{PAUSE} = \text{avg}(|pau_N(j) - pau_L(j)|) \quad (11)$$

4. 実験

4.1. 実験データ

3章で述べた評定手法を用いて自動評定値を算出し、教師評定値との相関関係を調べる実験を行った。実験データとして、「日本人による読み上げ英語音声データベース[8]」を使用した。英文は「韻律学習を考慮した読み上げ文セット」の「種々のイントネ

ーションに関する文」から選んだ10文である。実験に使用した英文を表1に示す。

表1 実験に使用した英文

Did you study Nelson?
That's from my brother, who lives in London.
I don't know him, personally.
Legumes are a good source of vitamins.
Have you locked a front door?
Yes. But did you see it yourself?
What do you need it for?
You left them at what time?
I haven't seen you before, have I?
Isn't it wonderful weather!

学習者は8名(男性)、発話比較のためのネイティブは1名(男性)である。学習者発話に対して、4名の英語教師(一部、3名)がイントネーションを5段階で評定している。英語教師は逐次、ネイティブ発話を聴取し、その発話を正解とした場合の評定を行っている[8]。英語教師4名の平均評定値を教師評定値とする。

なお、相関関係を調べるにあたり、教師が学習者発話のイントネーションを評定する際には韻律の三要素全ての情報を基に評定を行うと考え、教師によるイントネーション評定値と計算機による基本周波数、パワー、発話長それぞれの評定値との相関関係を調べる。

4.2. 実験結果

自動評定値と教師評定値との相関係数を表2に示す。

基本周波数に関しては、「基本周波数パターン距離」、「基本周波数パターン回帰直線の傾き差」を用いた評定手法において「単語境界部」を比較単位とすることの有効性が示されている。また、基本周波数とパワーに関して提案した「回帰曲線近似誤差の差」を用いた評定手法においては、「単語」を比較単位とすることである程度の相関を得た。発話長に関する結果からは、「単語長の差」を用いた評定手法が最も良い結果となり、発話時間とリズムの両方を組み合わせた評定が有効であることが判る。発話中の無音の長さを用いた評定手法では期待したような結果を得ることができなかった。

表2 自動評定値と教師評定値との相関

評定手法		比較単位	相関係数
基本周波数	パターン距離	発話全体	-0.29
		単語	-0.36
		境界シラブル	-0.47
		境界モーラ	-0.41
	回帰直線の傾き差	単語	-0.26
		境界シラブル	-0.36
		境界モーラ	-0.45
	回帰直線近似誤差の差	発話全体	-0.14
		単語	-0.37
		境界シラブル	-0.26
	回帰2次曲線近似誤差の差	発話全体	-0.15
		単語	-0.42
境界シラブル		-0.40	
パワー	回帰直線	発話全体	-0.18
	近似誤差の差	単語	-0.38
	回帰2次曲線	発話全体	-0.21
	近似誤差の差	単語	-0.27
発話長	総発話長の差		-0.25
	正規化単語長の差		-0.23
	正規化無音長の差		-0.02
	単語長の差		-0.44
	無音長の差		-0.13

4.3. 重回帰モデル

4.2節で示した実験結果を基に、教師評定値を推測するために有効な説明変数 X_n を選び重回帰分析を行った。重回帰分析に用いた説明変数を表3に、重回帰モデルを式(12)に示す。なお、式(12)は韻律パラメータ間の重要度を比較するために、各韻律パラメータを標準化(平均と分散を正規化)した後モデル化した結果である。

表3 重回帰分析に用いる説明変数

X_1 :	基本周波数パターン距離 (シラブル境界部)
X_2 :	基本周波数パターン回帰2次曲線 近似誤差の差(単語)
X_3 :	パワーパターン回帰直線近似誤差 の差(単語)
X_4 :	総発話長の差
X_5 :	正規化単語長の差

$$Y = -0.28X_1 - 0.18X_2 - 0.15X_3 - 0.22X_4 - 0.17X_5 + 2.93 \quad (12)$$

式(12)のモデルを評価するために、モデル作成時に用いた英文と同じ文セット中から別の発話データ(モデル作成時とは別の10文を男性8名が発声)を選びオープンデータとした。発話比較のためのネイティブはモデル作成時と同一話者、教師評定値も同一条件で収集されたデータである。モデルを評価した結果、相関係数は評価クロズドデータで0.70、オープンデータで0.57となった。

5. まとめと今後の課題

本研究では、単語境界部に注目した評定や「回帰曲線近似誤差の差」を用いた評定手法を提案し、教師評定値と強い相関を持つ評定値の算出を試みた。単語境界部に関しては2つの区間を検討し、どちらも単語や発話全体に注目するよりも良い結果を得ることができた。今後は単語境界部を統計的に決定する手法を検討し、自動評定値と教師評定値との相関をより強めたいと考えている。回帰曲線近似誤差の差を用いた評定手法については、回帰直線と回帰2次曲線以外の回帰曲線も用いて評定を行い、結果の違いを考察する必要がある。

また、今回の実験では発話比較に際して、ネイティブ話者1名の発話と学習者発話とを比較し評定値を算出した。今後は複数のネイティブ発話を用いることで、各比較単位における重要度を決定し重み付けして評定する手法や、最も標準的なネイティブ発話と学習者発話とを比較する手法を検討する予定である。

さらに、提案した評定手法や作成した韻律評定モデルを、「ネイティブ発話を聴いて真似した発話データ」に対しても適用し、有効性の有無を検証したいと考えている。

参考文献

- [1] 峯松信明, 中川聖一, “強勢音節検出技術に基づく英単語単発声からの発音上の癖推定”, 信学技報, SP2001-1, pp.9-16(2000).
- [2] 柳ヶ瀬裕則, 山下洋一, “発話比較による英語学習者の韻律自動評定”, 信学技報, SP2000-118, pp.9-16(2001.1).
- [3] 加藤圭介, 野沢和典, 山下洋一, “英語学習者発話における韻律自動評定方法の検討”, 音講論(2003.9).

- [4] 中川他, “音声認識技術を利用した英会話CAIシステム”, 情報処理学会論文誌, Vol.38, No8, pp.1649-1657(1997)
- [5] 杉藤美代子, “日本人の英語”, 和泉書院(1998).
- [6] 清水克正, “英語音声学”, 劉草書房(1995).
- [7] 小張他, “日本人の英語発音と英語母語話者音声韻律特徴の比較”, 音講論(2003.9).
- [8] 峯松他, “日本人英語音声に対する母語話者英語教師による評価ラベリング”, 音講論(2002.9).